Massachusetts Institute of Technology

Department of Electrical Engineering and Computer Science

6.242, Fall 2004: MODEL REDUCTION *

Models of dynamical systems[1]

The main goal of this lecture is to review the basic definitions of system theory, leading to the notion of order of an LTI model.

2.1 General systems and signals

While the main objects of study in this class are LTI systems of finite order, they will frequently be viewed in a context of interaction with other systems of other types. This section provides some minimal background in general systems and signals.

2.1.1 Signals and systems in continuous time

It is convenient to think of continuous time (CT) *signals* as real vector-valued functions of time $t \in (-\infty, \infty)$, integrable over every bounded interval $(-T, T)$, $T > 0$. From this viewpoint, $f_1(t) = |t|^{-1/2}$ (defined at zero by $f_1(0) = 0$) and $f_2(t) = e^{t^2}$ are signals, while $f_3(t) = 1/t$ (defined at zero by $f_3(0) = 0$) and $f_4(t) = \delta(t)$ (Dirac delta) are not. The set of all signals with values in $\mathbf{R}^k$ will be denoted by $\mathcal{L}^k$.

A continuous time *system* S with an m-dimensional input and k-dimensional output is simply a map from a subset $\mathcal{L}_0$ of $\mathcal{L}^m$ into $\mathcal{L}^k$ (usually multi-valued, so that one input $f \in \mathcal{L}_0 \subset \mathcal{L}^m$ corresponds to many possible outputs $y \in \mathcal{L}^k$). This definition reflects the fact that, in most dynamical models, a system's output is defined not only by the input but also by a set of auxiliary parameters called *initial conditions*, as in Figure 2.1. Hence

[1] Version of September 11, 2004

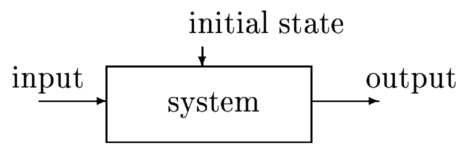

Figure 2.1: System with initial conditions

the output is not uniquely defined by the input.

For example, the familiar pure integrator system (transfer function $1/s$) maps a signal $f \in \mathcal{L}^1$ to signals of the form

$$y(t) = c_0 + \int_0^t f(\tau)d\tau,$$

where c_0 is an arbitrary constant playing the role of an initial state.

2.1.2 Signals and systems in discrete time

Let us view discrete time (DT) signals as continuous time signals which only change value at a discrete set of uniformly spaced time instances $t_k = kT$, where $T > 0$ is a fixed real number called the *sampling rate* of a DT signal, and k is a non-negative integer. The usual meaning of a discrete time signal is that at every time t it represents the *last available sample* of a continuous time signal, provided the samples are taken uniformly with interval T starting at zero time. For example, sampling CT signal $f(t) = \cos(\pi t)$ at rate $T = 1$ yields a discrete-time signal

$$f_d(t) = \begin{cases} 1, & k \leq t < k+1, \ k \in \{0, \pm 2, \pm 4, \dots\}, \\ -1, & k \leq t < k+1, \ k \in \{\pm 1, \pm 3, \pm 5, \dots\} \end{cases} .$$

Note that this f_d can also be viewed as a DT signal at sampling rate $1/M$ for every positive integer M (though, indeed, it will not be the result of sampling $f(t) = \cos(\pi t)$ at rate $T = 1/M$ for $M \neq 1$).

An alternative way of representing a discrete time signal $f = f(t)$ with sampling rate $T > 0$ is by specifying T and the sequence of its samples $f[k]$ at time instances $t_k = kT$, i.e.

$$f[k] = f(kT), \quad k = 0, \pm 1, \pm 2, \dots .$$

Thus, a DT signal $f(t)$ is completely defined by its sampling rate T and by the sequence of sampled values $f[k] = f(kT)$. For example, the DT signal $f_d(t)$ from above can be defined as such with sampling rate $T = 1$ and sampled values sequence $f_d[k] = (-1)^k$.

The set of all discrete time k-dimensional signals at sampling rate T will be denoted by $\mathcal{L}^k_{[T]}$.

A discrete time system S is a map from a subset $\mathcal{L}_0$ of $\mathcal{L}^m_{[T]}$ into $\mathcal{L}^k_{[T]}$ (usually multi-valued). Note that this definition requires same sampling rates for input and output. Therefore, a DT system with a k-dimensional input and m-dimensional output can also be viewed as a map $S: l^m_+ \mapsto l^k_+$, where l^q_+ denotes the set of all sequences $f = f[i]$ of q-dimensional real vectors, indexed by non-negative integers i. For example, the familiar one step delay system (transfer function $1/z$) maps $f = f[i]$ into $g[i] = f[i-1]$.

2.1.3 System state

The notion of a *system state* is very important for understanding dynamical system analysis and design. Informally speaking, system state at a given time t_0 is the information needed to define the output for $t \geq t_0$ when the input for $t \geq t_0$ is given, and to reconstruct the output for $t < t_0$ when the input for $t < t_0$ is known. Some system models have a "state vector" explicitly defined, but an adequate construction of a system state can also be given for every input/output description.

In general, what is actually needed is a notion of two input/output pairs (f_1, y_1) and (f_2, y_2) defining same state of a given system $\mathcal{S}$ at time T. Then a particular state of clS at time T is defined as a maximal set of input/output pairs defining same state at time T.

The formal definitions go as follows. Two input/output pairs (f_1, y_1) and (f_2, y_2) of system $\mathcal{S}$ are called *interchangeable* in $\mathcal{S}$ at time T if the "hybrids",

$$f_{12}(t) = \begin{cases} f_1(t), & t < T, \\ f_2(t), & t \geq T \end{cases}, \quad y_{12}(t) = \begin{cases} y_1(t), & t < T, \\ y_2(t), & t \geq T \end{cases},$$

and

$$f_{21}(t) = \begin{cases} f_2(t), & t < T, \\ f_1(t), & t \geq T \end{cases}, \quad y_{21}(t) = \begin{cases} y_2(t), & t < T, \\ y_1(t), & t \geq T \end{cases},$$

are also valid input/output pairs of $\mathcal{S}$. Two input/output pairs (f_1, y_1) and (f_2, y_2) of system $\mathcal{S}$ are said to *define same state* of $\mathcal{S}$ at time T if they are interchangeable in $\mathcal{S}$ at time T with the same set of input/output pairs.

For example, consider the system $\mathcal{S}$ for which the set $\mathcal{B}_\mathcal{S}$ of all possible input/output combinations (f, y) consists of all pais of scalar functions $f, y \in \mathcal{L}^1$ such that $e^t(y(t) - f(t))$ does not depend on time. It is easy to check that two pairs (f_1, y_1) and (f_2, y_2) are interchangeable in $\mathcal{S}$ at time T if and only if

$$y_2(T) - f_2(T) = y_1(T) - f_1(T).$$

Hence this condition is also necessary and sufficient for (f_1, y_1), (f_2, y_2) to define same state at time T. We can conclude that the state of an input/output pair (f, y) of $\mathcal{S}$ at time T can be associated with the real number

$$x(T) = y(T) - f(T).$$

One can even claim that the state defined this way evolves according to $\dot{x}(t) = -x(t)$, and thus produce a *state space model*

$$dx(t)/dt = -x(t), \quad y(t) = x(t) + f(t)$$

for the system.

As another example, consider a system $\mathcal{S}$ with no input and a scalar output $y[k] \in \{0, 1\}$ which is allowed to change not more than once. Here output $y_0[k] \equiv 0$ is interchangeable at time $T = 2$ with $y_1[k] = 1 - u[k]$, where

$$u[k] = \begin{cases} 0, & k < 0, \\ 1, & k \geq 0, \end{cases}$$

is the standard DT unit step function, and with $y_2[k] = u[k-5]$ but y_1 and y_2 are not interchangeable. Hence, one can see that interchangeability is not necessarily a transitive relation. It can be shown that the system allows a finite state automata model with a total of four states, which can be coded as pairs $(\theta[k], s[k])$, where $\theta[k] \in \{0, 1\}$ equals the current output value, $\theta[k] = y[k]$, and $s[k] \in \{0, 1\}$ equals 0 if and only if $y[m]$ has same value for all $m \in \{0, 1, \ldots, k\}$. The system equations can then be written in the em state space form

$$\theta[k+1] = \theta[k] + (1 - s[k])w[k], \quad s[k+1] = s[k] + (1 - s[k])w[k],$$

where $w[k]$ can be an arbitrary sequence of zeros and ones.

2.1.4 State space models

Examples from the previous subsection suggest that defining system models via equations of state evolution could be convenient. Indeed, state space models are known to be most useful in simulation and optimization of dynamical systems.

A *standard ODE state space model* of a CT system with m inputs, k outputs, and n states is defined by two functions $F: \mathbf{R}^n \times \mathbf{E}^m \mapsto \mathbf{R}^n$ and $G: \mathbf{R}^n \times \mathbf{E}^m \mapsto \mathbf{R}^k$ (possibly multi-valued), and declares an input/output pair (f, y) admissible if and only if there exists a solution $x = x(t)$ of the ODE

$$\dot{x}(t) = F(x(t), f(t)), \quad t \in \mathbf{R},$$

such that $y(t) = G(x(t), f(t))$ for all $t \in \mathbf{R}$. Note that some functions F may produce invalid models due to non-existence of solutions of the ODE. A number of extentions of the standard state space format are available.

In the CT example from the previous subsection, $n = m = k = 1$, and

$$F(\bar{x}, \bar{f}) = -\bar{x}, \quad G(\bar{x}, \bar{f}) = \bar{x} + \bar{f}.$$

A *standard difference equations state space model* of a DT system with m inputs, k outputs, and n states is defined by two functions $F: \mathbf{R}^n \times \mathbf{E}^m \mapsto \mathbf{R}^n$ and $G: \mathbf{R}^n \times \mathbf{E}^m \mapsto \mathbf{R}^k$ (possibly multi-valued), and declares an input/output pair (f, y) admissible if and only if there exists a solution $x = x[k]$ of the difference equation

$$x[k+1] = F(x[k], f[k]), \quad k \in \mathbf{Z} = \{0, \pm 1, \pm 2, \dots\},$$

such that $y[k] = G(x[k], f[k])$ for all $k \in \mathbf{Z}$.

It is easy to see that two input-output pairs (f_1, y_1) and (f_2, y_2) of the system defined by the state space model define same state at sample N whenever $x_1[N] = x_2[N]$. However, the opposite is not necessarily true, since different values of $x[N]$ do not have to result in a different prognosis for future/past. Practically, this is frequently a consequence of using redundant state components. For example, with $n = m = 1 = 1$,

$$F(\bar{x}, \bar{f}) = \bar{f}, \quad G(\bar{x}, \bar{f}) = \bar{f},$$

the future and the past do not depend on $x[N]$ at all.

2.2 LTI system models

This section discusses definitions and calculation of order for linear time-invariant (LTI) system models.

2.2.1 Causal CT convolution models

Let $h_c: [0, \infty) \mapsto \mathbf{R}^{k \times m}$ be an k-by-m matrix valued function which grows slower than a given exponent as $t \to \infty$, in the sense that

$$\int_0^\infty e^{-\sigma_h t} \|h_c(t)\| dt < \infty$$

for some $\sigma_h \geq 0$. Then $h_c(\cdot)$ defines a system which maps input signal $f \in \mathcal{L}^m$ with an exponentially bounded past, in the sense that

$$\sup_{t<0} e^{-\sigma_h t} |f(t)| < \infty,$$

into output signal $y \in \mathcal{L}^k$ defined by the convolution integral

$$y(t) = \int_0^\infty h_c(\tau) f(t-\tau) d\tau.$$

The resulting map $\mathcal{S}: \; f \mapsto y$ is causal, single-valued, linear, and time-invariant, in the sense that delaying $f(\cdot)$ by T units of time results in an identical delay of $y(\cdot)$.

The matrix function $h = h_c$ is called the *impulse response* of $\mathcal{S}$.

In general, it is convenient to allow $h = h(\cdot)$ to include a countable sum of delayed delta functions:

$$h(t) = h_c(t) + \sum_{r=0}^{\infty} h_k \delta(t - \tau_k),$$

where $\tau_k \geq 0$ and

$$\sum_{k=0}^{\infty} e^{-\sigma_h \tau_k} \|h_k\| < \infty,$$

in which case the input-output map is re-defined as

$$y(t) = \sum_{k=0}^{\infty} h_k f(t - \tau_k) + \int_0^\infty h_c(\tau) f(t-\tau) d\tau.$$

An associated *transfer matrix* of the LTI system is defined by

$$H(s) = \sum_{k=0}^{\infty} h_k e^{-\tau_k s} + \int_0^\infty h_c(\tau) e^{-\tau s} d\tau \quad (\mathrm{Re}(s) > \sigma_h).$$

When $\tilde{f} = \tilde{f}(s)$ is the Laplace transform of $f = f(t)$, the corresponding output $y = y(t)$ will have Laplace transform

$$\tilde{y}(s) = H(s) \tilde{f}(s) \quad \text{for} \quad \mathrm{Re}(s) > \sigma_h.$$

2.2.2 Finite Order State Space CT LTI Models

A state-space model for a finite order CT LTI system H with input $f = f(t)$, output $y = y(t)$, and state $x = x(t)$ has the form

$$\dot{x}(t) = Ax(t) + Bf(t), \tag{2.1}$$
$$y(t) = Cx(t) + Df(t), \tag{2.2}$$

where A, B, C, D are constant matrices with real entries.

$$H = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

frequently serves as a shortcut notation.

One common interpretation of (2.1),(2.2) associates it with an input/output model, in which a given input $f = f(t)$ prodices all outputs $y(t)$ which can be obtained by solving (2.1) first, with arbitrary initial conditions, and then defining y according to (2.1). This, however, is not compatible with convolution integral models, which define a unique output for a given input.

An alternative way of associating an input/output model with equations (2.1),(2.2) is given as follows. Assume that all eigenvalues of A have real part strictly smaller than $\sigma_A \geq 0$. Given an input $f = f(t)$ such that

$$\sup_{t<0} e^{-\sigma_A t}|f(t)| < \infty$$

the output $y = y(t)$ could be any signal for which there exists a function $x = x(t)$ such that

$$\sup_{t<0} e^{-\sigma_A t}|x(t)| < \infty$$

and equations (2.1),(2.2) are satisfied. A standard calculation shows that the output corresponding to $f = f(t)$ is given by

$$y(t) = Df(t) + \int_0^\infty Ce^{A\tau}Bf(t-\tau)d\tau.$$

Thus, the resulting input/output map acts as a convolution model with impulse response

$$h(t) = D\delta(t) + Ce^{At}Bu(t),$$

where $u = u(t)$ is the standard unit step function. The corresponding transfer matrix is given by

$$H(s) = D + C(sI - A)^{-1}B.$$

2.2.3 Order of a state space LTI model

It is tempting to refer to the dimension of $x(t)$ as the order of the corresponding state space model (2.1),(2.2). However, due to the posiible presence of redundant states, the true system order may be smaller.

A redundant state corresponds to the existence of a coordinate transformation which reduces the A matrix to an upper block triangular form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \tag{2.3}$$

while having

$$B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \tag{2.4}$$

or to the reducibility of A to the form

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \tag{2.5}$$

while having

$$C = \begin{bmatrix} C_1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \tag{2.6}$$

In both cases

$$H_r := \left(\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array} \right)$$

serves as an equivalent model with a reduced number of states.

A system for which such reduction is impossible called *minimal*. The following two theorems (offered here without proof) are helpful for understanding and dealing with minimality of LTI state space models.

The first theorem offers a simple criterion of minimality, equipped with an algorithm for reducing a non-minimal model.

Theorem 2.1 *Let n denote the dimension of the (square) matrix A.*

(a) Model (2.1),(2.2) is minimal if and only if matrices

$$M_c = [B \ AB \ A^2B \ \ldots \ A^{n-1}B]$$

and

$$M_o = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

have rank n.

(b) *If* $\text{rank}(M_c) < n$ *then there exists an invertible square matrix* S *such that*

$$SM_c = \begin{bmatrix} 0 \\ M_2 \end{bmatrix},$$

and the state change $Sx = x_{new}$ *transforms (2.1),(2.2) to a form (2.3),(2.4).*

(c) *If* $\text{rank}(M_o) < n$ *then*

$$M_o S = [0 \quad M]$$

for some invertible square matrix S*, and the state change* $x = Sx_{new}$ *transforms (2.1),(2.2) to a form (2.5),(2.6).*

A direct implementation of this theorem is likely to face severe numerical difficulties, because M_c (the controllability matrix) and M_o (the observability matrix) will be very poorly conditioned for large n. Nevertheless, numerically stable modifications of the algorithm are available, to be discussed later.

The second theorem explains the utility of minimal models in defining order of LTI system.

Theorem 2.2 *If two minimal models (2.1),(2.2) define same transfer matrix then the corresponding state vectors have an equal dimensions.*

In particular, order of a convolution system model can be defined as the number of states in an equivalent minimal finite order LTI state space model. The order equals infinity if such finite order model does not exist.

For example, the convolution system with impulse response

$$h(t) = \begin{bmatrix} e^{-t} & e^{-t} \\ e^{-t} & 0 \end{bmatrix}$$

and transfer matrix

$$H(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+1} \\ \frac{1}{s+1} & 0 \end{bmatrix}$$

has order 2. To see this, one can start with a state space model constructed by combining individual state space models for the three non-zero components of $H(s)$, as in

$$\begin{aligned} \dot{x}_1 &= -x_1 + f_1, \\ \dot{x}_2 &= -x_2 + f_1, \\ \dot{x}_3 &= -x_3 + f_2, \\ y_1 &= x_1 + x_3, \\ y_2 &= x_2, \end{aligned}$$

which is a state space model with

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

This model is not minimal, since the *controllability* matrix M_c satisfies $pM_c = 0$ where

$$p = [1 \quad -1 \ \ 0],$$

and the *observability* matrix M_o satisfies $M_o q = 0$, where

$$q = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

The row vector p indicates that $px = x_1 - x_2$ does not depend on the input (and hence identically equals zero in the "unique output" interpretation of the state space model). Therefore, it should be possible to reduce the original model by assuming that $x_1 = x_2$ at all times. Similarly, the column vector q indicates that the output does not change if a multiple of q is added to the system state $x(t)$. Therefore, it should also be possible to reduce the model by projecting $x(t)$ on a two-dimensional subspace in $\mathbf{R}^3$ along q.

One reduced mode is thus given by

$$\begin{aligned} \dot{x}_1 &= -x_1 + f_1, \\ \dot{x}_3 &= -x_3 + f_2, \\ y_1 &= x_1 + x_3, \\ y_2 &= x_1. \end{aligned}$$

A simple check shows that this model is minimal. Hence the order of the original convolution model equals 2.